

# Transferring Structures from PubChem to ACD/ChemSketch

---

Gavin Shear, Karim Kassam  
Advanced Chemistry Development, Inc.  
Toronto, ON, Canada  
[www.acdlabs.com](http://www.acdlabs.com)

## Introduction

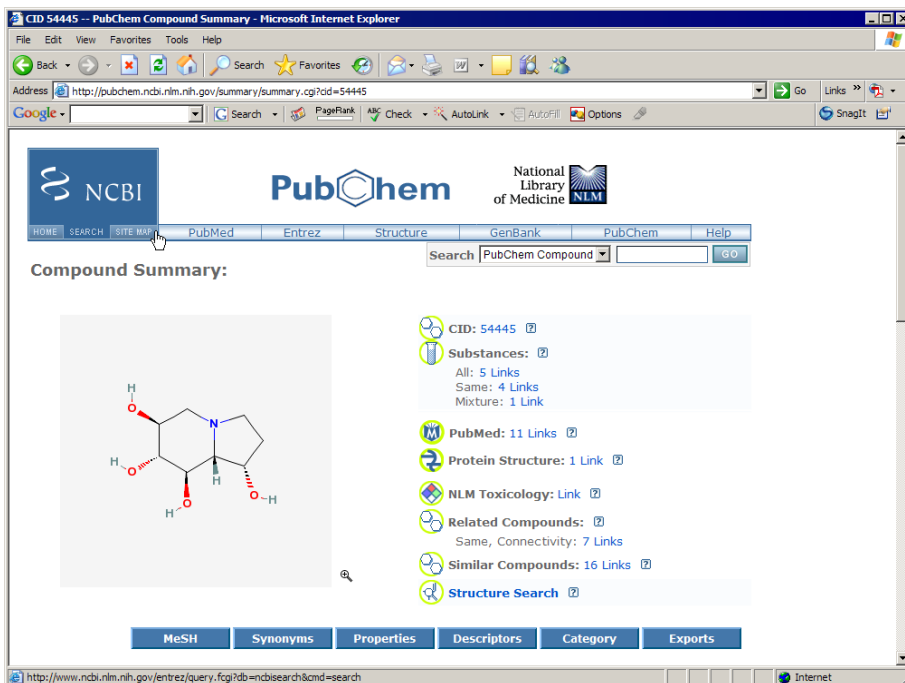
[PubChem](#) is an internet accessible database of small molecules made available to the public as part of the National Institute of Health's (NIH) [Molecular Libraries and Imaging](#) and [Roadmap Initiatives](#). The PubChem Web site offers extensive search capabilities and provides a variety of information on small molecules, including structural, physicochemical, and biological data. Structures can be readily transferred between the PubChem Web site and ACD/Chemsketch in a variety of ways.

Typical PubChem search results show the retrieved molecule along with common trade names, IUPAC name and InChI<sup>TM</sup> identifier, SMILES and canonical SMILES representations, as well as providing export capabilities to the SDfile format. ACD/ChemSketch not only makes it easy to accurately acquire compound data from PubChem, but it also offers convenient access to a variety of tools for manipulating, processing, or enriching the captured information with useful complementary scientific data.

Once the structure is brought into ACD/ChemSketch, a variety of useful chemical and analytical properties and inferences can be made using ACD/Labs modules to verify the IUPAC name provided, predict an NMR spectrum, predict log*P*, p*K*<sub>a</sub>, boiling point, and other physicochemical properties, and more. Further, the original data and extracted knowledge can now be incorporated into reports or pasted into Microsoft<sup>®</sup> Word documents.

This technical note describes two different approaches for transferring structure representation data from PubChem to ACD/ChemSketch. Structures from PubChem can be acquired either by using one of the text descriptors or by downloading structure data files. Each approach is equally reliable, so please feel free to choose whichever method best fits your work process and is most comfortable for you.

## Searching and Retrieving Molecules from PubChem



The screenshot shows a web browser window displaying the PubChem Compound Summary for CID 54445. The page includes the NCBI logo, the PubChem logo, and the National Library of Medicine logo. The main content area features a chemical structure of 1,2,3,5,6,7,8,8a-octahydroindolizine-1,6,7,8-tetrol. To the right of the structure, there is a list of related resources and links, including PubMed, Protein Structure, NLM Toxicology, Related Compounds, and Similar Compounds. The page also includes a search bar and navigation buttons for MeSH, Synonyms, Properties, Descriptors, Category, and Exports.

The PubChem site offers text-based searches according to chemical name, chemical property ranges, or elemental composition. In addition, PubChem allows substructure or structure similarity searches for drawn or imported structure files. Instructions on how to make advanced queries are provided at the PubChem site, at <http://pubchem.ncbi.nlm.nih.gov/search/>.

## IUPAC Name, SMILES, Canonical SMILES, and InChI™ Representations

In the **Descriptors Computed for Structure** section of the search-result page, the structure is provided in a variety of different text representations. As shown below, PubChem provides computed descriptors for a retrieved record using the IUPAC Name, Isomeric SMILES, Canonical SMILES, and InChI representations.

### Descriptors Computed from Structure: [?](#)

**IUPAC Name:** 1,2,3,5,6,7,8,8a-octahydroindolizine-1,6,7,8-tetrol  
**Isomeric SMILES:** C1CN2C[C@@H]([C@H]([C@@H]([C@H]2[C@H]1O)O)O)O  
**Canonical SMILES:** C1CN2CC(C(C2C1O)O)O  
**InChI:** InChI=1/C8H15NO4/c10-4-1-2-9-3-5(11)7(12)8(13)6(4)9/h4-8,10-13H,1-3H2/t4-,5-,6+,7+,8+/m0/s1

It is straightforward to copy and paste textual descriptors for molecules into ACD/ChemSketch in order to generate structure drawings usable by other ACD/Labs modules. ACD/ChemSketch provides the ability to correctly and unambiguously generate structures from text representation, thereby avoiding costly errors due to inaccurate transcriptions or incorrect structure drawing translation.

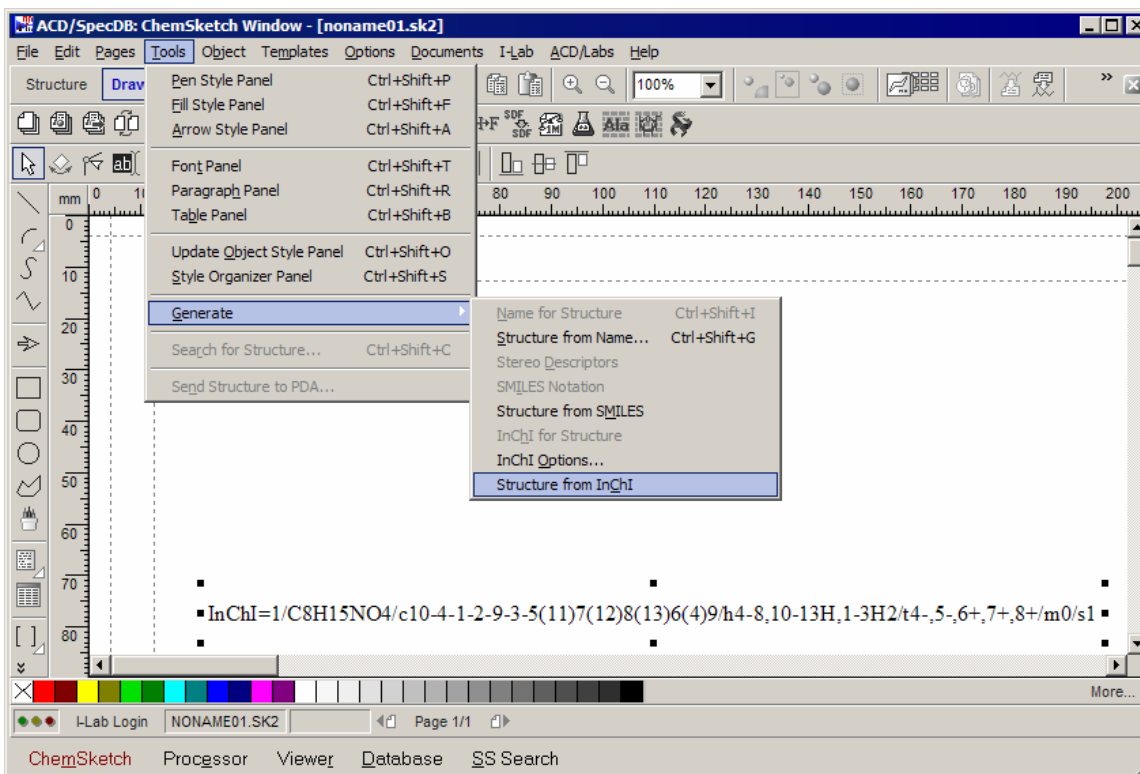
## Copying and Pasting the Structure Representation to ACD/ChemSketch

To generate a chemical drawing from a structure representation in PubChem, take the following steps:

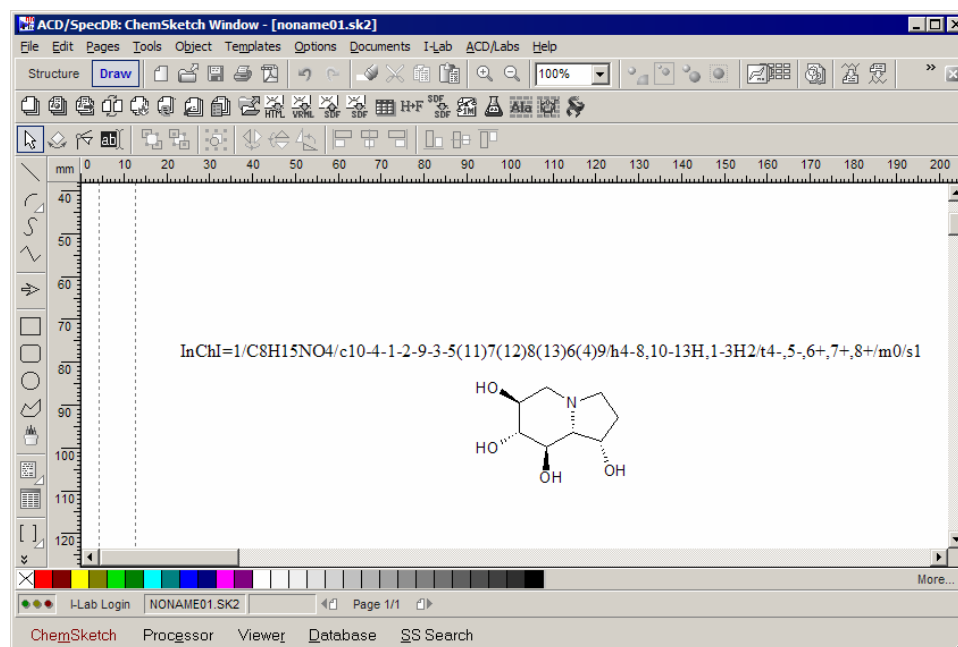
1. Select the text for either the IUPAC™ name, SMILES, or InChI™ descriptors and then press CTRL-C to copy the information to the Clipboard.

**Note** Make sure the entire string is selected as it may run for two or more lines.

2. Switch to the ACD/ChemSketch application, and then press CTRL-V to paste the structure text to its proper location in the ACD/ChemSketch windowpane. The text should now be visible and should still be selected, if not, click the text object to re-select it.
3. From the **Tools** menu, point to **Generate**, and then click the proper structure representation, i.e., **Structure from InChI** or **Structure from SMILES** or **Structure from Name**.



If required, you can edit the structure generated in ACD/ChemSketch. For example, you can use the Clean Structure tool or adjust atom positions manually. You can use color, change the font and bond size, apply custom atom numbering, and perform a number of other stylistic changes to improve the appearance and to highlight interesting scientific features. Another important feature is the ability to apply specific journal styles to make high-quality and publication-ready structure drawings. Further, you can paste structures from ACD/ChemSketch directly to Microsoft Word or other report applications.



Remember that ACD/ChemSketch is integrated directly into many powerful ACD/Labs modules, and that the structure can now be actively managed and leveraged. For example, the structure can be used as input for industry standard predictive and analytical modules to generate calculated NMR spectra or physicochemical properties. In addition, the structure can be pushed to a state-of-the-art searchable structure management system to create your own custom database (requires ACD/ChemFolder) or to supplement acquired analytical data (ACD/SpecManager).

## Downloading Structure Files from PubChem

At the bottom of each PubChem compound summary Web page are links to download the structure in a number of formats: ASN1, XML, and SDF. The SDfile (structure-data file) is probably the most widely used and recognized format for describing molecular structures. ACD/Labs' database modules can import SDfiles directly, so that the chemical structure and associated chemical information provided by PubChem are readily available.

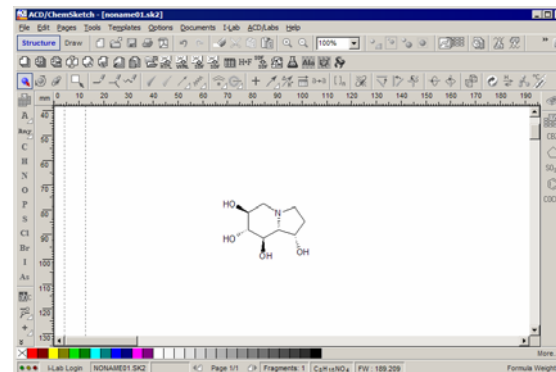
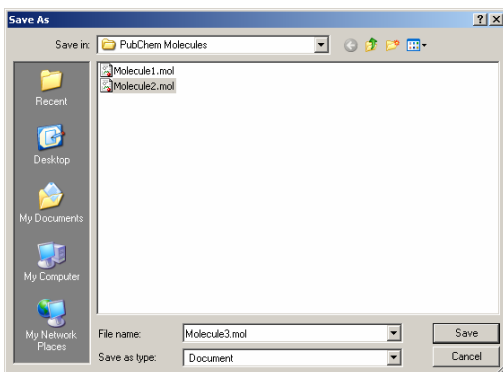


To import just the chemical structure into ACD/ChemSketch for an SDfile, you can simply apply a ".mol" extension to an SDfile. The program will then recognize the first structure in the record as the relevant chemical structure and ignore additional data and structures.

**Note** An SDFfile is essentially a series of molfiles with additional data fields for associated properties and information. Do not use the molfile extension if the molecule of interest isn't the first one in the SDFfile. You may use the SDF-to-Sketch converter available in the Goodies or the database module if you need to import all the structures contained in the SDfile.

To import a structure in SDfile format that was exported from PubChem, take the following steps:

1. At the bottom right of the PubChem page, click the **Save SDF** button.
2. In the 'File Name' box of the 'Save As' dialog box, type in the desired file name followed by the extension ".mol".
3. Under the **File** menu in ChemSketch, use the **Import** command to import the document created in the previous step.



## Conclusion

The ability to easily and accurately transfer structures from the PubChem database, as well as other sources into ACD/ChemSketch provides increased access to chemical information. The chances of costly errors due to inaccurate structure translations are greatly reduced. As well, the fact that chemical information can easily be managed, exploited, and distributed via ChemSketch and other ACD/Labs software is conducive to improved productivity. Once structures are in the ACD/ChemSketch interface, they can be utilized readily by the multiple prediction and databasing tools offered by ACD/Labs as well as with other applications such as other cheminformatic systems and common reporting applications like Microsoft Word or Adobe® PDF.